



Should the impact factor of the year of publication or the last available one be used when evaluating scientists?

Gustavo A. Slafer (Slafer, GA)^{1,2} and Roxana Savin (Savin, R)¹

¹ University of Lleida-AGROTECNIO Center, Dept. of Crop and Forest Sciences. Av. R. Roure 191, 25198 Lleida, Spain. ² Catalanian Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

Abstract

Aim of study: A common procedure when evaluating scientists is considering the journal's quartile of impact factors (within a category), many times considering the quartile in the year of publication instead of the last available ranking. We tested whether the extra work involved in considering the quartiles of each particular year is justified

Area of study: Europe

Material and methods: we retrieved information from all papers published in 2008-2012 by researchers of AGROTECNIO, a centre focused in a range of agri-food subjects. Then, we validated the results observed for AGROTECNIO against five other European independent research centres: Technical University of Madrid (UPM) and the Universities of Nottingham (UK), Copenhagen (Denmark), Helsinki (Finland), and Bologna (Italy).

Main results: The relationship between the actual impact of the papers and the impact factor quartile of a journal within its category was not clear, although for evaluations based on recently published papers there might not be much better indicators. We found unnecessary to determine the rank of the journal for the year of publication as the outcome of the evaluation using the last available rank was virtually the same.

Research highlights: We confirmed that the journal quality reflects only vaguely the quality of the papers, and reported for the first time evidences that using the journal rank from the particular year that papers were published represents an unnecessary effort and therefore evaluation should be done simply considering the last available rank.

Additional keywords: citation; paper impact; scientist evaluation; journal quality

Abbreviations used: IF (Impact Factor); JCR (Journal Citation Reports); NIF (Normalised Impact Factors); PI (Principal Investigator)

Authors' contributions: GAS: conceived and designed the analysis; contributed data; performed analyses. RS: contributed data; revised the analyses. Both authors wrote and approved the final manuscript.

Citation: Slafer, GA; Savin, R (2020). Should the impact factor of the year of publication or the last available one be used when evaluating scientists? Spanish Journal of Agricultural Research, Volume 18, Issue 3, eM01. <https://doi.org/10.5424/sjar/2020183-16399>

Supplementary material: (Table S1 and Fig. S1) accompanies the paper on SJAR's website

Received: 16 Jan 2020. **Accepted:** 21 Dec 2020.

Copyright © 2020 INIA. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC-by 4.0) License.

Funding: The authors received no specific funding for this work

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to Gustavo A. Slafer: slafer@pvcf.udl.cat

Introduction

Scientists are evaluated almost continuously for their scientific achievements/merits. In general, the works published in scientific journals are the core of such evaluation. This is because the ultimate aim of science is to generate new knowledge, and unless this knowledge has been published in a rigorous journal¹ it is unlikely to be considered seriously by any other scientist or science

administrator. The rationale is that only publication in such journals enables the new knowledge to be recognised and available to the rest of the world, including the author's peers who will then confirm or challenge the conclusions. Thus, the publication of new knowledge in recognised scientific journals is the foundational source of scientific knowledge.

Consequently, published papers provide the strongest credit for evaluation of the capability of a scientist to

¹ A journal warranting strong rigor in the acceptance of manuscripts for publication, mainly based on the originality and relevance of the tested hypotheses as judged by a strong and thorough peer-review system.

produce new and valuable knowledge, provided that the meaning of authorship is not devalued (Slafer, 2005; Rajasekaran *et al.*, 2014; Logan *et al.*, 2017). Although it might be ideal that true experts in the field review each paper to assign value to the contributions made by evaluated scientists, there are serious limitations for this when evaluating several researchers simultaneously (the most common scenario of evaluation) (Kreiman & Maunsell, 2011), and when researchers of different areas are evaluated together by a single panel. The time required for expert review of a relevant number of papers of a number of scientists would be impractical, and when a large number of experts are involved with each one evaluating only a few papers there is a serious bias produced by the inherent subjectivity, making the outputs of different peer reviewers (each evaluating different scientists) barely comparable. Consequently, it has been customary in evaluation processes to use quantitative tools to gauge the relative performance of evaluated people (particularly in recent years; Ancaiani *et al.*, 2015), even though reducing the assessment to a simple number might be dangerous (Sahel, 2011; Egghe, 2011). The first and simplest has been productivity: *i.e.* simply the number of published scientific papers, assuming that the greater the number of papers published the larger the overall contribution to knowledge.

However, papers vary hugely in their relevance (in general as well as within their specific field of knowledge) as effective advancers of knowledge (Abramo *et al.*, 2019). Even though quantity and quality are not necessarily at odds with each other (*e.g.* van Raan, 2013; Huang, 2016), it has been argued many times that focusing the evaluation on the quantity of papers, regardless of their quality, may not only be a poor measure (essentially because it does not take into account the importance of the papers; Hirsch, 2005) but may also send the wrong message to researchers who might feel inclined to ignore the quality of the journals in which they publish in pursuit of increases in productivity (*e.g.* Butler, 2002). Conversely, the opposite may be true when the evaluation focus switches from quantity to quality (*e.g.* Moed, 2008). Many attempts have been made to assess the quality of the scientific production of a scientist, mostly based on the number of citations (Waltman, 2016). The most successful of these has been the h-index developed by Jorge E. Hirsch (2005), which soon after its publication began to be a common tool used globally to quantify scientific research output, harmonising quantity and quality very simply. A traditional way to discriminate the quality of papers has been to assume that this is reflected by the quality of the journal in which they are published; journals within each particular field of knowledge are known to vary enormously in their prestige and importance. As predicted by Bradford in the 1930s, a small proportion of journals account for a large proportion of what is well-regarded by the community

(Bradford, 1934). Even though there is an overall poor relationship between the impacts of the individual papers and the impact factor of the journals in which they are published, owing to the fact that the journal impact factor reflects the average of a highly skewed distribution of impacts of individual papers (*e.g.* Seglen, 1992; 1997; Leydesdorff, 2008; Slafer, 2008; Mutz & Daniel, 2012), this relationship improves markedly if the analysis is restricted to journals within the same field (Slafer, 2008). In turn, this is the basis for using “normalised impact factors” when comparing scientists across disciplines (Owlia *et al.*, 2011; Bornmann *et al.*, 2013). This is consistent with the fact that the impact factor of the journal seems relevant for predicting the citation impact of published papers, particularly for recently published ones (Levitt & Thelwall, 2008; Abramo *et al.*, 2010; Didegah & Thelwall, 2013; Vancley, 2013; Stegehuis *et al.*, 2015) and that the impact factor of the journal may be positively associated with peer-reviewed scores to the journals (Liu *et al.*, 2015). Thus, the quality of the journal publishing the article is frequently used as a simple indicator of the presumed quality of the paper (Huang, 2016), which again, while being far from accurate, is practical when analysing recently published papers and within particular fields of knowledge (*e.g.* Slafer, 2008; Huang, 2016). Thus, using the quality of the journal to assess the quality of the research in the papers published is a widely adapted practice (see discussion in Chavarro *et al.*, 2018).

A procedure that has been commonly used consisted of categorising the journals within a particular field of knowledge into four quartiles (Q1-Q4 in decreasing order of impact; *i.e.* Q1 = journals within the top 25% of impact factors within their categories) and assign a value to each particular paper that is inversely proportional to the quartile in which it belongs (*i.e.* the lower the Q the higher the value of the paper). In particular, this has been applied to assess the likely impact of recently published papers, whose small number of citations may be scarcely meaningful. Additionally, recent publications may be more important than historic ones when the future performance of scientists needs to be assessed (Bornmann *et al.*, 2013), which is the case in the vast majority of evaluations. In the practicalities of the process, evaluators are frequently required to compute the value assigned to a paper by considering the quartile (or the impact factor) of the journal in the year that the paper was published. This means that if we consider the publications in the last 4 years we need to compute four (potentially) different values for papers published in the same journal. This extra work of finding and computing the quartile and impact factor of the journals for each particular year (instead of simply computing the last available figures) would only be reasonable if there were (i) a solid rationale for it, and (ii) empiric evidence that it has a significant impact on the result of the evaluation. As evaluators we have been in

this situation ourselves a number of times and we never received a solid explanation to justify the extra burden of needing to consider different impact factors/quartiles to evaluate the presumed quality of papers published in the same journal over the previous few years.

During a specific call in 2013 we were responsible for collecting and analysing a significant amount of information on productivity and impact for the period 2008-2013 of our host institution, AGROTECNIO (Center for Research in Agrotechnology), which is within the CERCA Centres (the Catalan research centres of excellence). AGROTECNIO is an interesting case for a bibliometric study because it hosts research groups that work on a relatively diverse range of disciplines across crop, environmental, animal, food and nutrition sciences. We publish in journals belonging to several different research categories. We thus found ourselves in a position where we could engage in a parallel study to test empirically whether a significant difference exists between evaluating the quartile of the journal in the year of publication and using the last available rank instead. After analysing the data from AGROTECNIO we expanded the work to include data within the same categories of research from other universities in Europe to validate the conclusions.

Material and methods

Due to a specific call in 2014, AGROTECNIO was required to prepare a detailed analysis of its publications over the five-year-period 2008-2012 (inclusive) at a time when the 2013 version of Journal Citation Reports (JCR) was the latest one available. For this purpose, we retrieved information from all papers co-authored by researchers of AGROTECNIO. There were 759 retrieved papers published in 257 different journals (Table S1 [suppl.]) belonging to 45 different subject categories in JCR 2013 (when a journal was included in more than one category, we selected the category closest to the subject matter of the paper), although c. 75% of the papers of the Center were published in journals categorised in seven categories (AGRICULTURE DAIRY & ANIMAL SCIENCE, AGRICULTURE MULTIDISCIPLINARY, AGRONOMY, ENTOMOLOGY, FOOD SCIENCE & TECHNOLOGY, PLANT SCIENCES, and VETERINARY SCIENCES), which is consistent with the main focus of AGROTECNIO's research agenda. Later we analysed with the Web of Science (accessed in December 2014) and its associated Journal Citation Reports (i) the number of citations received by each of the papers and (ii) the rank of the journal within its scientific category.

With the number of citations received by each paper and its age (time since publication and citation counting) we calculated the mean citation rate per year for each article. For this exercise, we considered all citations

(including self-citations). We did so not only because it is the most common procedure in evaluations (particularly when the number of people being evaluated is large), but also and mainly because self-citations are expected to result from of a cohesive research program, in which authors must refer to their previous papers to justify subsequent contributions to knowledge (Cooke & Donaldson, 2014), and can be considered equally important as cites from others (Kacem *et al.*, 2020). In high-standard journals it is expected that reviewers and editors judge, among many other things, that the authors used the most relevant references and, in that context, it should be assumed in principle that self-citation may not be simply a misconduct (that naturally may also be the case; Bartneck & Kokkelmans, 2011; Ioannidis, 2015).

Using the rank of the journals in each category we classified them among of the four quartiles, with Q1 being the top rank that comprised journals with the highest impact factors in their categories and Q4 being the bottom quartile that included journals with the lowest impact factors. We have carried out this classification using both the JCR of the year in which the paper was published (the year-of-publication JCR) and the JCR of 2013 (the last available version of JCR at the time we retrieved all the information). Journal impact factors considered were the traditional measure of the average number of citations received in a particular year by articles published in the last two years, as published by JCR (*i.e.* citations in year y of items published in years $y-1$ and $y-2$ divided by the number of citable items published in years $y-1$ and $y-2$).

By adding up the values of each of the papers published by each Principal Investigator (PI) in the period considered we then obtained values (and rankings) corresponding to each PI using either the year-of-publication JCR for each paper or JCR 2013 across all the papers. To mirror the procedure frequently followed in real evaluations we assigned a value to each paper that was inversely proportional to the quartile in which it belonged. To make this simple we assigned 4 points to papers in journals ranked within the top quartile (Q1), 3 points to papers in Q2-journals, 2 points to those in Q3-journals and finally 1 point to papers published in journals ranked in the fourth quartile (Q4). In addition, for each individual paper we assigned this value by considering the quartile from the year in which the paper was published (which is the common practice; in this case JCRs 2008, 2009, 2010, 2011, and 2012), and also considered the quartile corresponding to the last available JCR (in this case JCR 2013). The scientific structure of AGROTECNIO at the time of the exercise included 85 researchers (including staff, postdocs and PhD students) divided into 13 research groups of different sizes, each of them headed by a PI. To test whether it is better to use the particular year of publication or just the last available JCR to assign a particular value to each paper, we selected these PIs and assessed

their publications for the period 2008-2012, as explained above.

Because the level of performance of an individual centre may be unexpectedly skewed, we validated all the results observed for AGROTECNIO against five other independent research centres with prestigious reputations in AGROTECNIO fields of interest (the entire food chain including crop, animal, environmental, and food sciences, as well as related aspects like soil sciences and nutrition; and covering fundamental and applied aspects). For this validation exercise, from Spain we selected the Technical University of Madrid (UPM) and internationally the Universities of Nottingham (UK), Copenhagen (Denmark), Helsinki (Finland), and Bologna (Italy). For each of these centres we retrieved equivalent information and made the same calculations mentioned above. For this purpose, we accessed the Web of Science – Core Collection, at the same time as the AGROTECNIO search, and retrieved all papers published from these within the same core scientific categories identified for AGROTECNIO: AGRICULTURE DAIRY & ANIMAL SCIENCE, AGRICULTURE MULTIDISCIPLINARY, AGRONOMY, ENTOMOLOGY, FOOD SCIENCE & TECHNOLOGY, PLANT SCIENCES, and VETERINARY SCIENCES (Table S1 [suppl.]). With these retrieved data we made the same calculations described for AGROTECNIO. Because we did not know the PIs at these universities we selected six researchers from each, three being the most frequent senior authors and the other three being the most frequent last authors of the retrieved papers. Once we selected the six scientists from each of the five institutions we listed all their papers (regardless of their position in the by-lines) and as described above for the AGROTECNIO PIs, we assigned values to each paper that were inversely proportional to the quartile of the journal in which it was published by considering either the year-of-publication JCR or JCR 2013.

For analysing the data, we had into account that the distribution of cites per paper and year were not normal, exhibiting a large degree of heteroscedasticity (and that is why the average number of cites per paper and year was much smaller than the mean between the most and least cited paper in each quartile). To cope with this issue we carried out ANOVAs with data transformed using the root square of the variable (as there were 0 cites per paper and year we could not correct with logarithms). After running the ANOVAs we verified the correct validation of the model through plotting the residuals of the transformed variable and verifying that heteroscedasticity disappeared. To analyse the relationships between the impact factor of the specific years evaluated (2008-2012) and the impact factor of the last available year (2013) we used linear regression. In all cases where we fitted this relationship we verified the correct validation of the model as well.

Results

Productivity and quality of database analysed

The AGROTECNIO database analysed not only varied across a range of categories (Table S1 [suppl.]) but also represents an international centre with a reasonable scientific standard. Not only was the productivity more than reasonable (with an average of more than 150 papers published in SCI-indexed journals per year for a relatively small centre) but also the quality of the journals in which the papers were published was of a high standard (Fig. 1, left panel).

Analysis of the specific impacts of individual papers from AGROTECNIO (number of citations received by each paper divided by the time since publication and when the data were collected) indicated that not only were they mainly published in the highest impacting journals within each research field, but the impacts of the individual papers were also high in general. More than half of the papers (median of 2 cites paper⁻¹ year⁻¹) had normalised impact factors higher than 1 (NIFs, the ratio of citations received by a paper to the global average citations per paper in the same field, 1 = world average; Langfeldt *et al.*, 2015). And the average paper of AGROTECNIO had an impact 50% higher than globally expected in the same

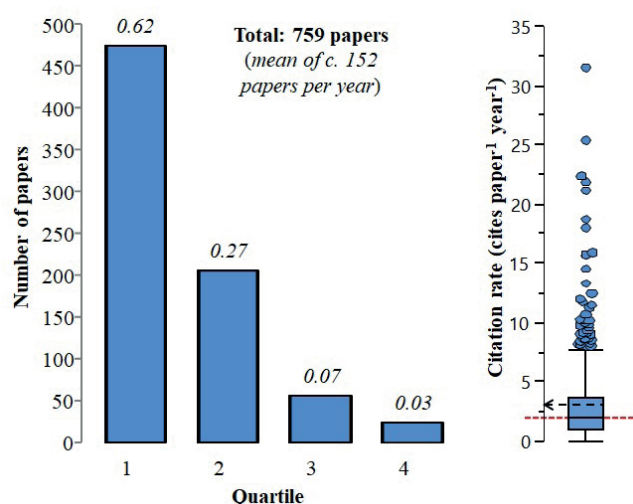


Figure 1. Left panel: Number of papers published by AGROTECNIO researchers in the period 2008-2012 in SCI-indexed journals belonging to the different quartiles when ranked within their categories of the Journal Citation Reports of 2013. Figures above each bar represent the proportion of papers in each quartile. Right panel: boxplot showing the distribution of the annual citation rates of the 759 papers published. The black dashed arrow indicates the mean annual citation rate of all AGROTECNIO papers, while the red dashed line shows the global average for the last ten years (2008-2017) for the Research fields of AGRICULTURAL SCIENCES and PLANT & ANIMAL SCIENCE using the baselines provided in the essential indicators of the Web of Science

field of knowledge (*i.e.* had a NIF of 1.52; Fig. 1, right panel). Furthermore, there were a large number of papers that had attained remarkable annual citation rates.

Does the journal quality reflect the quality of the papers?

As mentioned above, in evaluation systems that focus on recently published papers, it is common to assign a presumed value to a paper based on the value of the journal. To make this simpler (if not directly achievable when a large number of scientists must be evaluated) and comparable across field categories, the quartile of the journal in which the papers are published (directly visible in the Web of Science) is used instead of the impact factor of the journal.

The relationship between the actual impact of the papers (*i.e.* the annual citation rate of each paper) and the impact factor quartile of a journal within its category was

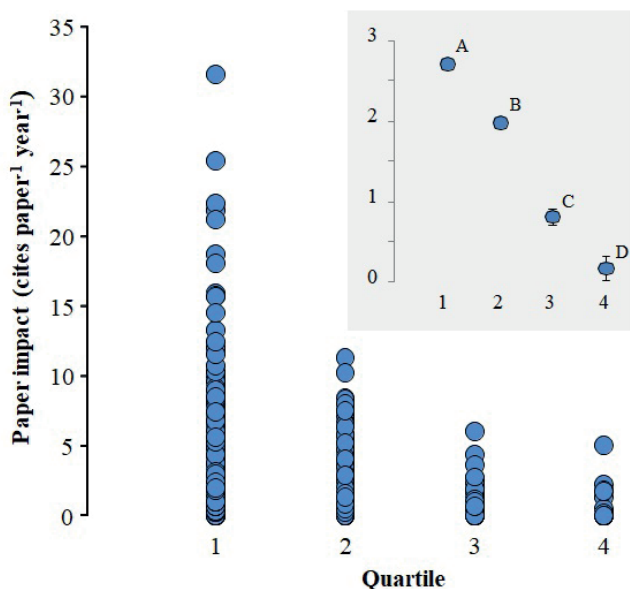


Figure 2. Impact of each individual paper published by AGROTECNIO researchers in 2008-2012 and the journal publication quartile within its category (rankings from Journal Citation Reports 2013). The main figure shows each of the 759 papers plotted individually and the inset is the average paper impact (and the corresponding standard error, whose magnitude was smaller than the size of the symbol if not seen) for all papers published in all journals belonging to the same quartile (averages and standard errors are the outputs of the ANOVA with the paper impact transformed using the root square of the cites per paper and year). Different letters on top of each average indicate that the averages were significantly different following a Tukey's honest significance test (HSD).

not clear (Fig. 2). The situation did not improve when we estimated the quality of the journal as a continuous variable, using the impact factor percentile (Fig. S1 [suppl.]), rather than as the discrete division of four quality classes following the four quartiles.

There was indeed a trend for papers published by AGROTECNIO researchers in low quartile journals to have less impact than those published in top quartile journals, but the proportion of the variation in the papers' impacts that was explained by the quality of the journal was very low (Fig. 2; Fig. S1 [suppl.]). This was due to the fact that the degree of variation in the impact of papers published in journals of the same quartile was very large. Nevertheless, it was also true that the degree of variation was larger in the top quartiles than the bottom quartiles (as can be seen by the spread of data-points for each quartile in Fig. 2, and the resulting magnitude of the standard errors of the means in the inset). Therefore, even though it is true that very low impact papers were found in journals ranked in any of the four quartiles, it was only in the top ranked journals that there were papers with very high actual impact (Fig. 2; Fig. S1 [suppl.]). In other words, the quality of the journal did not relate to the impact achieved by the least impacting papers, but it was a fair reflection of the impact of successful papers. These results are not just a peculiarity of AGROTECNIO researchers. Exactly the same patterns² were seen in the other four European institutions we selected for benchmarking/complementing the AGROTECNIO analysis (Fig. 3).

Furthermore, the likelihood of finding papers with very low impact in any of the four quartiles was strongly related to the quality of the journal (Fig. 4). Indeed, the proportion of papers published in journals ranked in the top quartile of their category that had no citations in the few years following publication was less than 2%, and this was also very low for papers published in journals of the second quartile (approximately 3%), but the likelihood rose noticeably to more than 15% in Q3 journals and reached a worrying 62% in Q4 journals (Fig. 4). On the other hand, when using an annual citation rate of two citations per year, which is a reasonable standard in the fields of knowledge embraced by AGROTECNIO researchers, the likelihood of reaching at least this level was very high in papers published in Q1 journals and it decreased noticeably when the papers were published in higher quartile journals. Even considering the likelihood of an average of eight or more citations per year (an extremely large citation rate in the mentioned fields), close to 10% of the papers published in Q1 journals reached this standard, while the proportion diminished to a third of

² Strictly speaking, there was a slight departure from the general pattern in the case of the University of Helsinki, as there was a paper with a very high citation rate published in a Q2 journal. That exception is a paper published in *Phytotaxa*, a journal ranked in Q2 (in the JCR of 2013), which is by far the most cited paper in the history of this journal (almost doubling the number of citations received by the second most quoted paper; which is in turn also another of the data-points with a very high citation rate in the same figure)

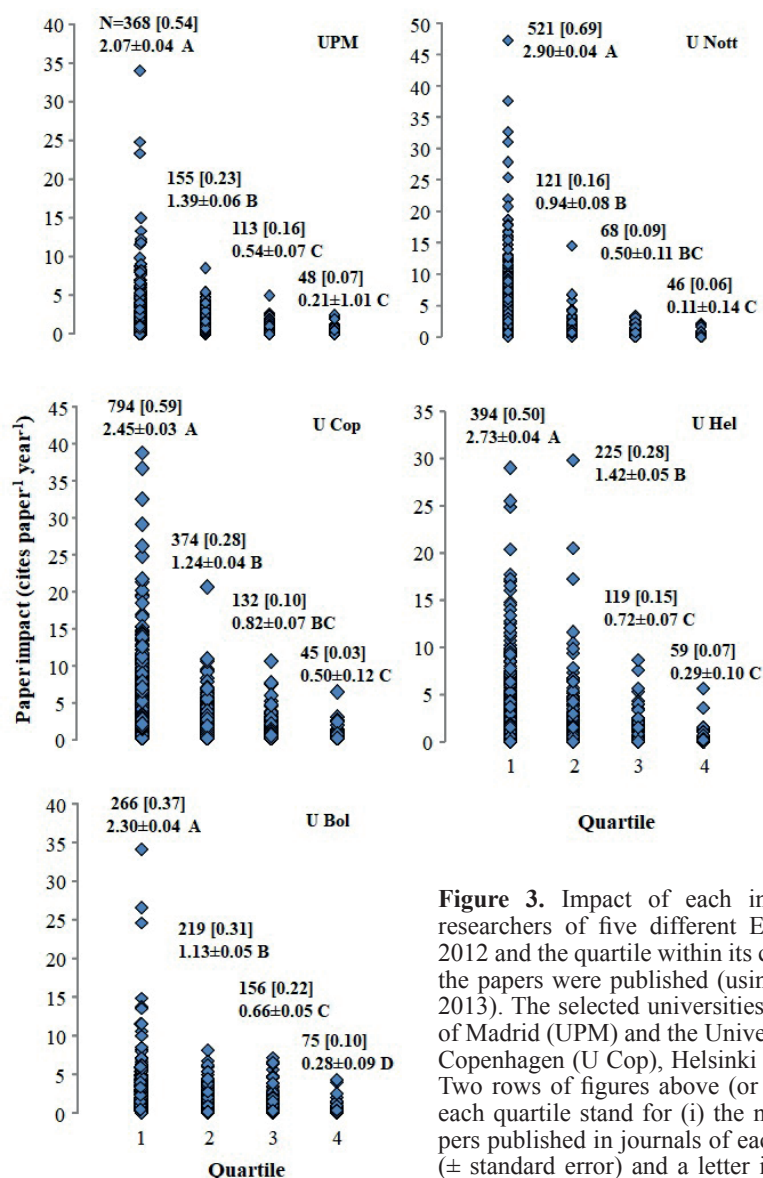


Figure 3. Impact of each individual paper published by researchers of five different European universities in 2008-2012 and the quartile within its category of the journal in which the papers were published (using the Journal Citation Reports 2013). The selected universities were the Technical University of Madrid (UPM) and the Universities of Nottingham (U Nott), Copenhagen (U Cop), Helsinki (U Hel), and Bologna (U Bol). Two rows of figures above (or on the right) of data-points of each quartile stand for (i) the number [and proportion] of papers published in journals of each quartile, and (ii) the average (\pm standard error) and a letter indicating, when different, that they were significantly different ($p=0.05$) following a Tukey's honest significance test (HSD); averages standard errors and significance taken from the ANOVA done with data transformed (square root of cites per paper and year).

that value in Q2 journals, and no papers had reached this level of excellence in terms of impact in journals of the last two quartiles (Fig. 4).

Do we need to consider the specific year of a paper's publication to assess its quality?

Assuming then that in some circumstances paper quality must follow the quality of the journal in which it is published (a rather widespread practice), should we consider the quality of the journal in the year of publication? Or, could we skip the extra work required to look for journal rankings in each year and simply use the last available rank? Answering these questions is not trivial, given the time it could take to evaluate each journal rank for the

number of years under consideration. The basis for this requirement is that the impact factor of journals changes from year to year. Even though in the majority of cases these changes are small, they may bring about changes in their relative rankings (which may also change by inclusions/exclusions of journals in the category considered).

We analysed this question for the publication database derived from AGROTECNIO by comparing the impact factor of the journal (encompassing with a very large range of journals; Table S1 [suppl.]) in the year in which the paper was published (*i.e.* 2008, 2009, 2010, 2011, or 2012) and the last available impact factor at the time of analysis (2013). All papers but four were published in journals ranging in impact factor from 0.1 to approximately 10-15. The four "outliers" were papers published in Nature Biotechnology (impact factor in

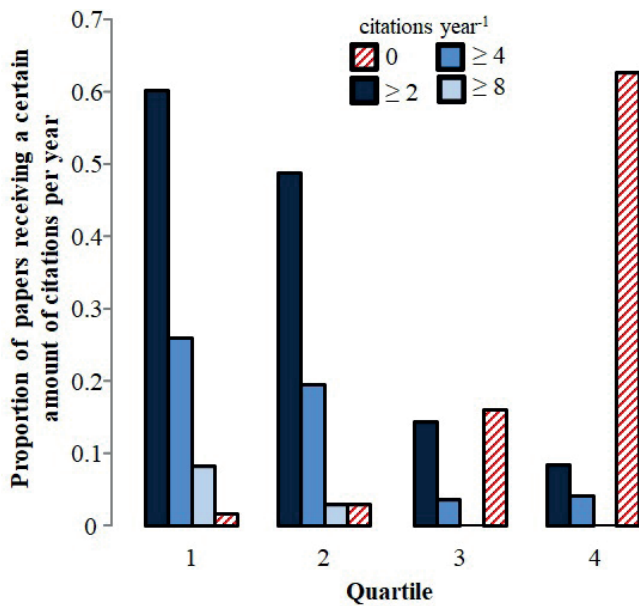


Figure 4. Proportion of papers receiving no citations at all (red-striped bars) or receiving at least 2 (dark blue bars), 4 (blue bars), or 8 (light blue bars) citations per year in journals ranked in the different quartiles of their categories

2013 of almost 40). For this analysis we omitted these four papers to ensure a homogeneous distribution of impact factors, avoiding the whole set of relationships being heavily skewed by a single case.

It was clear that the impact factor for each journal differed from year to year. Although there was an overall high degree of concordance between journal impact factors as reported in JCR 2013 and during the five preceding years, the relationship was not perfect (Fig. 5). In fact, the relationship was very strong (with coefficients of determination and regression close to 1) when the difference between the JCR issues was only one year; but both coefficients tended to decrease with an increase in the difference between the JCRs used to determine the IF of the journal (Fig. 5). But even when considering the relationship with the largest difference in years in the analyses done, the 2013 impact factor of the 256 journals explained more than 85% of the variation in their 2008 impact factors (Fig. 5).

These relationships do not represent a particularity of the journals in which AGROTECNIO researchers published. Validating the output of this analysis, the same trends were also found for the publication outputs of the other centres investigated (Fig. 6).

While the differences between the journal impact factors over a relatively short interval of years were generally minor, as expected, they might still have affected the outcome of the evaluation due to the difference between the journal impact factors immediately above and below the

quartile thresholds also being minor. Therefore, we decided to test directly whether and how much the outcome of the evaluation of researchers would be affected by ignoring these minor differences by using the journal quartiles from the last available JCR.

Does this variation in IF across years alter the rankings between the scientists being evaluated?

In order to answer this question, we graded each of the AGROTECNIO PIs exclusively by the accumulated value of their publications for the period 2008-2012, awarding each of their publications 1, 2, 3 or 4 points for papers published in Q4, Q3, Q2 or Q1 journals, respectively. When we compared the grades achieved by each PI relative to the journal quartiles in the particular year of publication or the last available journal rank there was almost no difference: the coefficient of regression was very close to 1, the intercept very close to the origin and the coefficient of determination close to 1 (Fig. 7, left panel). Indeed, the data-points fell very close to the line representing the 1:1 ratio.

Because differences in the values would be smaller than the corresponding differences in the rankings, we analysed the relationships in the rankings as well. Although the differences became somewhat more visible, the coefficients of regression and determination were still close to 1 and the intercept was also close to the origin (Fig. 7, right panel).

Again to confirm that this lack of substantial differences between using year-of-publication quartiles versus last-available quartiles was not unique to our centre's researchers, we included another 30 researchers in the analysis, 6 from each of the 5 European universities selected. Even though the degrees of freedom increased from 12 to 42, the coefficient of determination remained very close to 1, and almost without exception all the data points clustered on the 1:1 ratio line (Fig. 8).

Discussion

The quality and quantity of research produced by AGROTECNIO in the period analysed, when compared to the small size of the centre, reveals that it is a high-standard research institute and therefore makes the conclusions of this exercise trustworthy. This confidence in the conclusions is further warranted by the validation made with results from five other large independent universities with prestigious reputations in the same disciplinary field in Europe (one from Spain the other four from the UK, Denmark, Italy, and Finland).

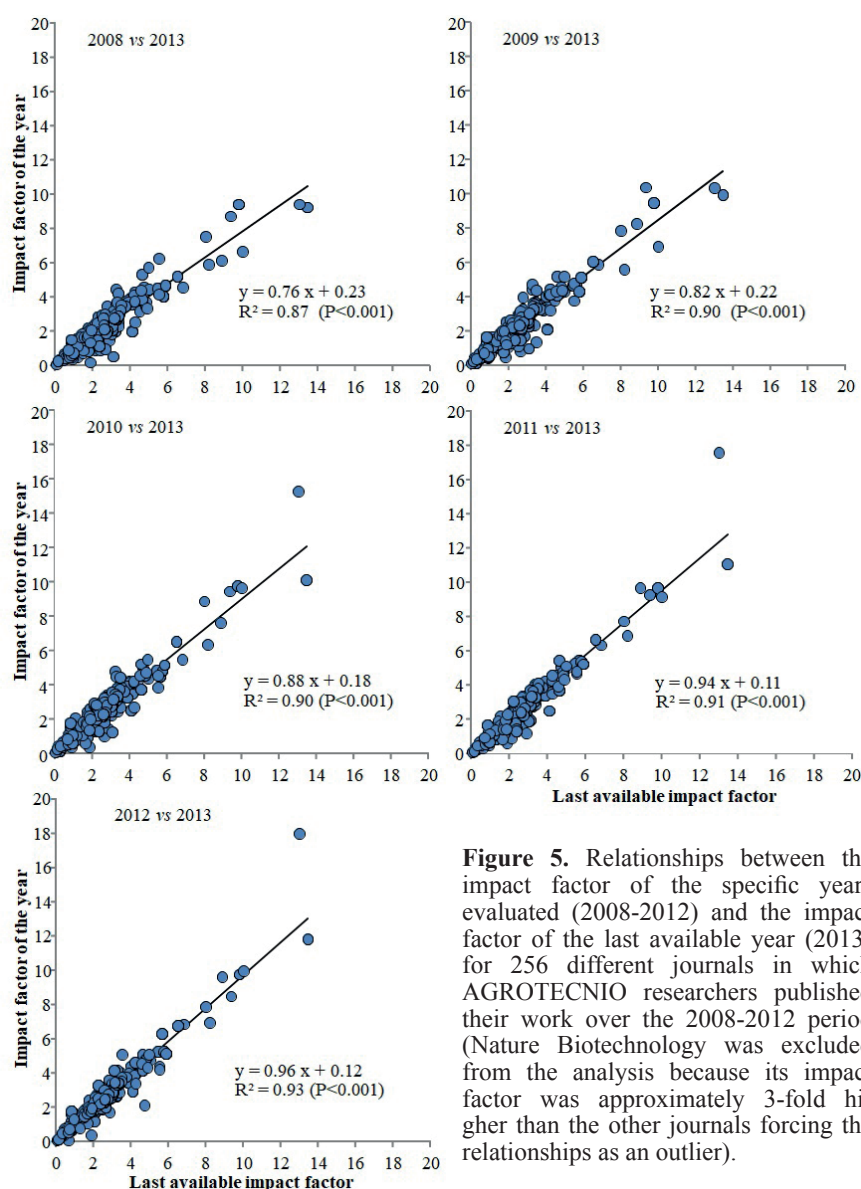


Figure 5. Relationships between the impact factor of the specific years evaluated (2008-2012) and the impact factor of the last available year (2013) for 256 different journals in which AGROTECNIO researchers published their work over the 2008-2012 period (Nature Biotechnology was excluded from the analysis because its impact factor was approximately 3-fold higher than the other journals forcing the relationships as an outlier).

Does the journal quality reflect the quality of the papers?

If it does, it would not be an entirely accurate reflection! Indeed, it would be rather naïve to expect an accurate estimate of the impact of an individual paper via the average impact of all papers in that journal (which is behind the impact factor of the journal), knowing that the distribution of citation rates is not Gaussian. Indeed, the impact factor of a journal is highly determined by the impact of a relatively low proportion of all papers published (Seglen, 1997; Frank, 2003; Slafer, 2008). The skewness of citations seems to be a generalised pattern (*e.g.* Bornmann & Leydesdorff, 2017) indicating that a relatively small percentage of the papers published in a journal (say *c.* 10%) account for a significant share (say *c.* 50%) of all citations received by the journal and that a large percentage of papers are uncited (Seglen, 1992; Albarra

et al., 2011). Consequently, it cannot be assumed that the impact factor of the journal can be used to accurately estimate the impact of most of the papers it publishes.

Having said that, it is also true for recently published papers that we have very few clues about their actual quality beyond the journal in which they were published. In support of the idea of using the journal quality as a proxy when nothing else is available, in the present study we found a clear trend in the papers' average citation rates and the quartile to which the journals belong, and this trend seems more than trivial. We demonstrated this relationship not only for AGROTECNIO's researchers but also for those of another five universities across Europe with excellence in plant, animal and food sciences. Furthermore, this clear trend is commensurate with the large number of cases reporting that the journal's impact factor may be relevant to the impact that a paper has in its field (*e.g.* Levitt & Thelwall, 2008; Abramo *et al.*, 2010; Didegah & Thelwall, 2013; Vanclay, 2013; Stegehuis *et al.*, 2015).

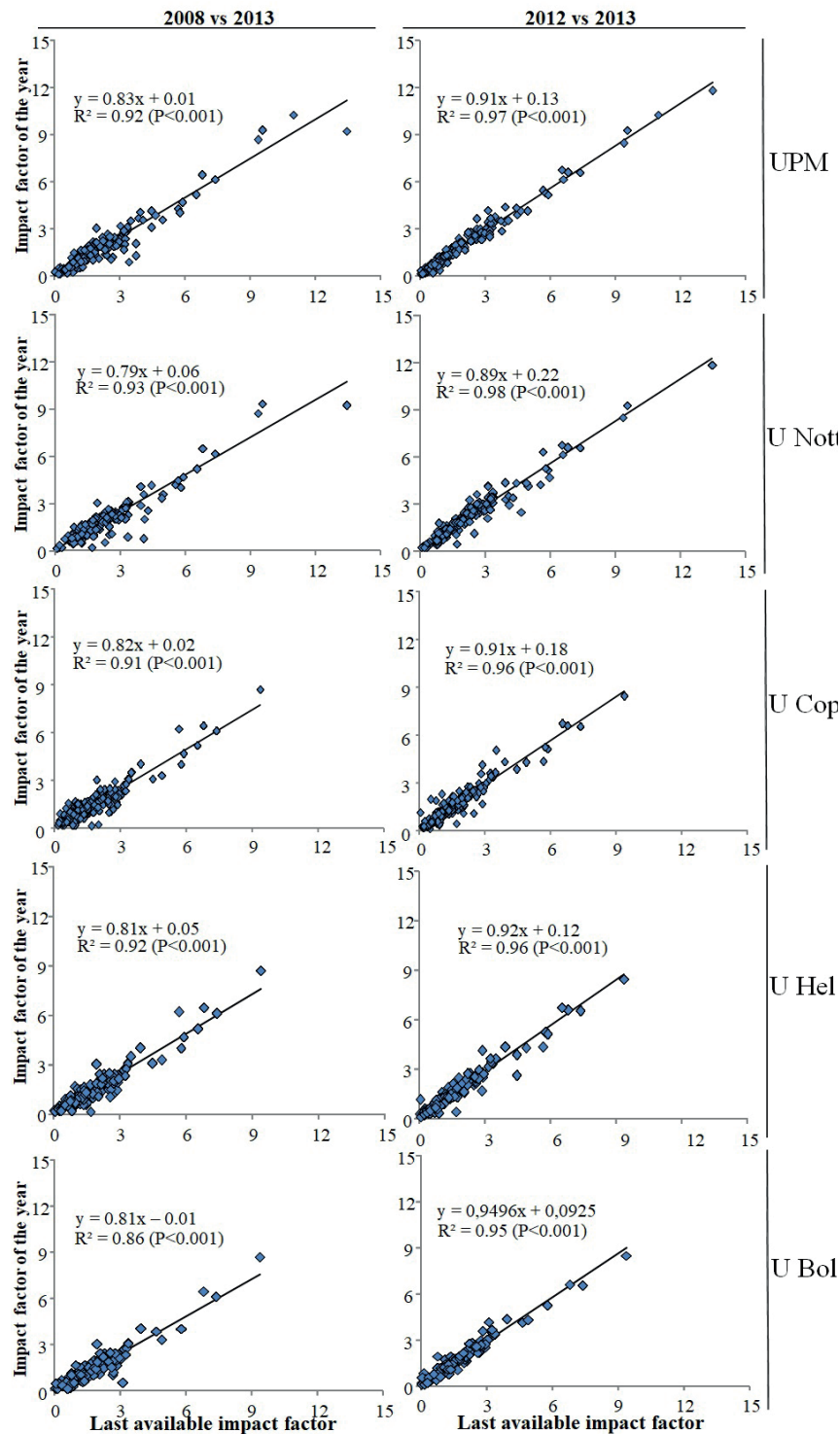


Figure 6. Relationships between the impact factors of the earliest (left panels) and latest (right panels) years of the publications considered (2008 and 2012, respectively) and the last available impact factor when we analysed the data (2013) for journals in which researchers of the Technical University of Madrid (UPM), and the Universities of Nottingham (U Nott), Copenhagen (U Cop), Helsinki (U Hel), and Bologna (U Bol) had published in the same core areas of research as AGROTECNIO. In the case of the Univ. of Nottingham, one journal (*Annual Review of Plant Biology*) was excluded from the analysis because its impact factor (18.9) was clearly higher than the other journals, and as an outlier would have skewed the relationships

In addition, there are other intuitive (though hard to neglect) arguments in favour of accepting that the quality of the journal is an indication of the quality of the papers.

Indeed, authors are more inclined to submit what they consider their best papers to highest-ranked journals in their fields and, on top of that, these more prestigious

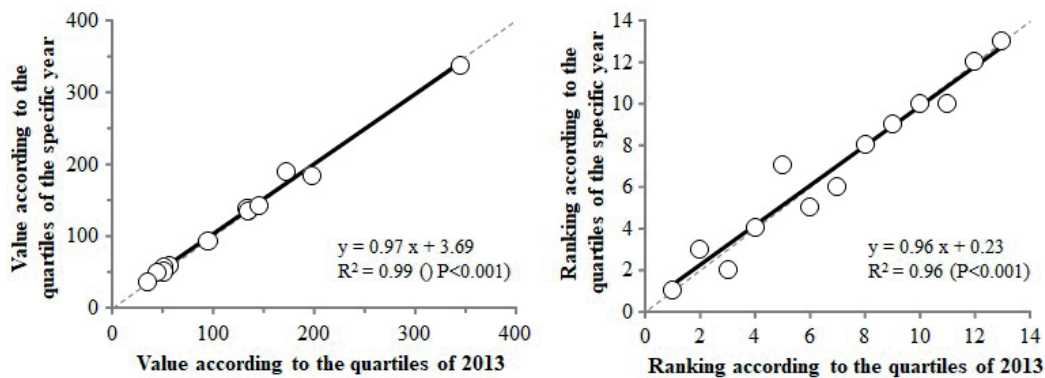


Figure 7. Relationship between values (left panel) or rankings (right panel) assigned to each of the 13 AGROTECNIO PIs based on their publications in 2008-2012 according to the JCR journal quartiles of the year of publication vs JCR 2013 (the last available JCR at the time of analysis). To assign values in this case we applied a linear increase in value per paper of 1 for papers in Q4, through to 2 (Q3), 3 (Q2) and 4 for papers in Q1. Dashed and solid lines represent the 1:1 ratio and the regression line, respectively

journals receive more submissions, which in turn enable editor opinion to exert strong discrimination towards the journal's acceptance of the best manuscripts for publication. In addition, in times when relying on high-quality peer reviewers is becoming more difficult (*e.g.* Baveye & Trevors, 2011; Fox, 2017) and engaging reliable peer reviewers has been lately rather frustrating for editors in general (and a true nightmare for some individual manuscripts), more prestigious journals tend to have less

difficulty in recruiting the best reviewers than less prestigious journals.

All in all, and even if far from perfect, we believe that using the quality of the journal as a rough proxy for the likely impact of the paper is acceptable when (i) the evaluation has a requirement to focus on recently published work, and (ii) involves a large enough cohort of scientists that would make it impossible to read each of the published papers to establish a subjective expert opinion³. Paraphrasing the famous phrase from Churchill about democracy, assuming the quality of recently published papers by the quality of the journal in which they are published may be the worst form of evaluation, except for all those other forms that have been tried from time to time.

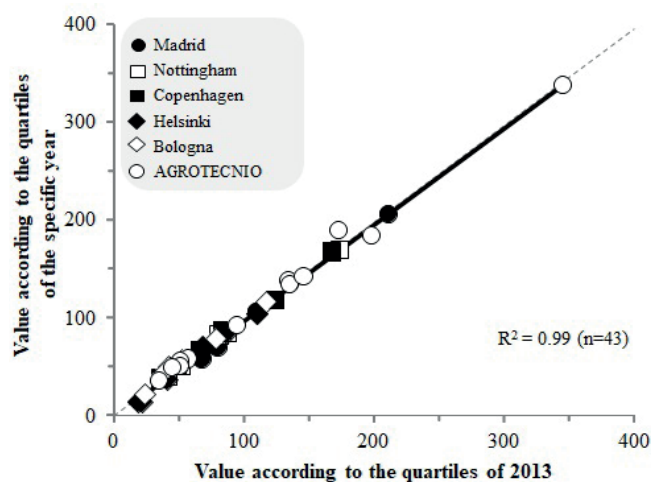


Figure 8. Relationship between values assigned to each of the 6 researchers selected from each of the five European Universities, together with those of the 13 AGROTECNIO PIs, based on their publications in 2008-2012 assigning the quartile for the journal according to the year-of-publication JCR for each paper vs JCR 2013 (the last available JCR at the time of analysis). Dashed and solid lines represent the 1:1 ratio and the regression line, respectively

Is it necessary to determine the journal rank for the year of publication?

Despite the fact that the impact factors of the journals change from year to year, it seems totally unnecessary to determine the rank of the journal for each year under analysis when using the last available rank at the time of evaluation for all published years considered produces an almost identical outcome (always considering the evaluation of relatively recently published work, for longer periods the situation may well be different; *e.g.* Pajić, 2015). Focusing only on the last available rank of journals saves a large amount of the evaluators' time when assigning a particular value to each contribution, without significant consequences for the evaluation output. As evaluators undertake their job by stealing time from their professional

³ Experts are recognised as such because they are very active scientists at the frontier of knowledge of their field: by definition they have very limited time and it will therefore be impossible for a large number of experts to dedicate a huge amount of time to an evaluation process. The contribution of experts in evaluation, even if not reading the papers in detail, is still essential because they can identify many types of misconduct (duplications, salami papers, inappropriate assigning of authorship, and so on) that if not detected would result in credit rather than penalties for the offenders.

activity or personal life (in most cases rather generously), the feeling that they are wasting time plays strongly against the likelihood of engaging them in the process.

Although we are not aware of any other empirical analyses similar to the current work, our conclusion is in line with the report of Finardi (2013), who analysed the evolution over time of impact factors and mean received citations. Finardi (2013) concluded that it does not make sense to use the JCR of the specific year of publication because there is no systematic change that improves/decreases the quality of a journal in such a short window of time.

Furthermore, even in the hypothetical case where there might be solid reasons to accept that the last available rank cannot provide an unbiased estimate of paper quality relative to annual rankings in the preceding years, why should the year of publication be used? There is no way that an individual paper could have any influence on the impact factor of the journal in the year it was published. Using the rank of the journal in the year of publication would mean assigning a particular value to a paper given as an average of the papers published in the previous two years. If there were any reason not to use the last available rank of the journal, the average of the two years following the year of publication should be used instead. This is because it is only over the two years following publication that the evaluation of specific papers affects the impact factor and the rank of the journal. This in fact would mean that there would be no reference at all for gauging papers published during the two years immediately before the evaluation process! Fortunately, as our empirical analysis has shown and the fact that there is no systematic change in journal quality over short periods of time (a few years) it seems valid to use the last available rank to assign a presumed value to papers published in the preceding years.

All in all, we conclude that the practice of using the journal rank from the particular year that papers were published when evaluating recent scientific output should be considered an unnecessary investment of time for evaluation that should be avoided, and instead it is recommended to simply use the last available journal rank.

Acknowledgement

We do gratefully acknowledge the advice on statistical analyses by Prof. Ignacio Romagosa (Univ of Lleida).

References

- Abramo G, D'Angelo CA, Di Costa F, 2010. Citations versus journal impact factor as proxy of quality: Could the latter ever be preferable? *Scientometrics* 84: 821-833. <https://doi.org/10.1007/s11192-010-0200-1>
- Abramo G, D'Angelo CA, Felici G, 2019. Predicting publication long-term impact through a combination of early citations and journal impact factor. *J Informetrics* 1: 32-49. <https://doi.org/10.1016/j.joi.2018.11.003>
- Albarrán P, Crespo J, Ortuño I, Ruiz-Castillo J, 2011. The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics* 88: 385-397. <https://doi.org/10.1007/s11192-011-0407-9>
- Ancaiani A, Anfossi AF, Barbara A, Benedetto S, Blasi B, Carletti V, Cicero T, Ciolfi A, Costa F, Colizza G, *et al.* 2015. Evaluating scientific research in Italy: The 2004-10 research evaluation exercise. *Res Eval* 24: 242-255. <https://doi.org/10.1093/reseval/rvv008>
- Bartneck C, Kokkermans S, 2011. Detecting h-index manipulation through self-citation analysis. *Scientometrics* 87: 85-98. <https://doi.org/10.1007/s11192-010-0306-5>
- Baveye PC, Trevors JT, 2011. How can we encourage peer-reviewing? *Water Air Soil Pollut* 214: 1-3. <https://doi.org/10.1007/s11270-010-0355-7>
- Bornmann L, Leydesdorff L, 2017. Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data. *J Informetrics* 11: 164-175. <https://doi.org/10.1016/j.joi.2016.12.001>
- Bornmann L, Leydesdorff L, Wang J, 2013. Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-rank approach (p100). *J Informetrics* 7: 933-944. <https://doi.org/10.1016/j.joi.2013.09.003>
- Bradford SC, 1934. Sources of information on specific subjects. *Engineering* 137: 85-86.
- Butler L, 2002. A list of published papers is no measure of value. *Nature* 419: 877. <https://doi.org/10.1038/419877a>
- Chavarro D, Ràfols I, Tang P, 2018. To what extent is inclusion in the Web of Science an indicator of journal 'quality'? *Res Eval* 27: 106-118. <https://doi.org/10.1093/reseval/rvy001>
- Cooke S, Donaldson M, 2014. Self-citation by researchers: Narcissism or an inevitable outcome of a cohesive and sustained research program? *Ideas Ecol Evol* 7: 1-2. <https://doi.org/10.4033/iee2014.7.1.e>
- Didegah F, Thelwall M, 2013. Determinants of research citation impact in nanoscience and nanotechnology. *J Am Soc Inform Sci Technol* 64: 1055-1064. <https://doi.org/10.1002/asi.22806>
- Egghe L, 2011. A disadvantage of h-type indices for comparing the citation impact of two researchers. *Res Eval* 20: 341-346. <https://doi.org/10.3152/095820211X13164389670356>
- Finardi U, 2013. Correlation between journal impact factor and citation performance: an experimental study. *J Informetrics* 7: 357-370. <https://doi.org/10.1016/j.joi.2012.12.004>

- Fox CW, 2017. Difficulty of recruiting reviewers predicts review scores and editorial decisions at six journals of ecology and evolution. *Scientometrics* 113: 465-477. <https://doi.org/10.1007/s11192-017-2489-5>
- Frank M, 2003. IFs: arbiter of excellence? *J Medical Library Assoc* 91: 4-6.
- Hirsch JE, 2005. An index to quantify an individual's scientific research output. *P Nat Acad Sci USA* 102: 16569-16572. <https://doi.org/10.1073/pnas.0507655102>
- Huang DW, 2016. Positive correlation between quality and quantity in academic journals, *J Informetrics* 10: 329-335. <https://doi.org/10.1016/j.joi.2016.02.002>
- Ioannidis JPA, 2015. A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *J Psychosomatic Res* 78: 7-11. <https://doi.org/10.1016/j.jpsychores.2014.11.008>
- Kacem A, Flatt JW, Mayr P, 2020. Tracking self-citations in academic publishing. *Scientometrics* 123: 1157-1165. <https://doi.org/10.1007/s11192-020-03413-9>
- Kreiman G, Maunsell JH, 2011. Nine criteria for a measure of scientific output. *Front Comput Neurosci* 5: 48. <https://doi.org/10.3389/fncom.2011.00048>
- Langfeldt L, Bloch C, Sivertsen G, 2015. Options and limitations in measuring the impact of research grants - evidence from Denmark and Norway. *Res Eval* 24: 256-270. <https://doi.org/10.1093/reseval/rvv012>
- Levitt JM, Thelwall M, 2008. Patterns of annual citation of highly cited articles and the prediction of their citation ranking: A comparison across subjects. *Scientometrics* 77: 41-60. <https://doi.org/10.1007/s11192-007-1946-y>
- Leydesdorff L, 2008. Caveats for the use of citation indicators in research and journal evaluations. *J Am Soc Inform Sci Technol* 59: 278-287. <https://doi.org/10.1002/asi.20743>
- Liu XL, Gai SS, Zhang SL, Wang P, 2015. An analysis of peer-reviewed scores and impact factors with different citation time windows: A case study of 28 ophthalmologic journals. *PLoS ONE* 10 (8): e0135583. <https://doi.org/10.1371/journal.pone.0135583>
- Logan JM, Bean SB, Myers AE, 2017. Author contributions to ecological publications: What does it mean to be an author in modern ecological research? *PLoS ONE* 12 (6): e0179956. <https://doi.org/10.1371/journal.pone.0179956>
- Moed HF, 2008. UK research assessment exercises: Informed judgments on research quality or quantity? *Scientometrics* 74: 153-161. <https://doi.org/10.1007/s11192-008-0108-1>
- Mutz R, Daniel HD, 2012. Skewed citation distribution and bias factor: Solutions to two core problems with the journal impact factor. *J Informetrics* 6: 169-176. <https://doi.org/10.1016/j.joi.2011.12.006>
- Owlia P, Vasei M, Goliaei B, Nassiri I, 2011. Normalized impact factor (NIF): An adjusted method for calculating the citation rate of biomedical journals. *J Biomed Informatics* 44: 216-220. <https://doi.org/10.1016/j.jbi.2010.11.002>
- Pajić D, 2015. On the stability of citation-based journal rankings. *J Informetrics* 9: 990-1006. <https://doi.org/10.1016/j.joi.2015.08.005>
- Rajasekaran S, Shan RLP, Finnoff JT, 2014. Honorary authorship: Frequency and associated factors in physical medicine and rehabilitation research articles. *Archiv Phys Med Rehabil* 95: 418-428. <https://doi.org/10.1016/j.apmr.2013.09.024>
- Sahel JA, 2011. Quality versus quantity: Assessing individual research performance. *Sci Transl Med* 3: 84cm13. <https://doi.org/10.1126/scitranslmed.3002249>
- Seglen PO, 1992. The skewness of science. *J Am Soc Inform Sci* 43: 628-638. [https://doi.org/10.1002/\(SICI\)1097-4571\(199210\)43:9<628::AID-ASIS>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASIS>3.0.CO;2-0)
- Seglen PO, 1997. Why the IF of journals should not be used for evaluating research. *Brit Med J* 314: 497-502. <https://doi.org/10.1136/bmj.314.7079.497>
- Slafer GA, 2005. Multiple authorship of crop science papers: are there too many co-authors? *Field Crops Res* 94: 272-276. <https://doi.org/10.1016/j.fcr.2004.11.011>
- Slafer GA, 2008. Should crop scientists consider a journal's impact factor in deciding where to publish? *Eur J Agron* 29: 208-212. <https://doi.org/10.1016/j.eja.2008.07.001>
- Steghuis C, Litvak N, Waltman L, 2015. Predicting the long-term citation impact of recent publications. *J Informetrics* 9: 642-657. <https://doi.org/10.1016/j.joi.2015.06.005>
- Van Raan AFJ, 2013. Universities scale like cities. *PLoS ONE* 8: e59384. <https://doi.org/10.1371/journal.pone.0059384>
- Vanclay JK, 2013. Factors affecting citation rates in environmental science. *J Informetrics* 7: 265-271. <https://doi.org/10.1016/j.joi.2012.11.009>
- Waltman L, 2016. A review of the literature on citation impact indicators. *J Informetrics* 10: 365-391. <https://doi.org/10.1016/j.joi.2016.02.007>